

Paul Paul Cloarec–Pioffet et Baptiste Rivoirard



**ADSP Free Project : Modification de la composition
cellulaire et transcriptomique du tissu musculaire en
réponse à l'exercice physique**

Introduction

Le séquençage d'ARN en cellule unique (scRNA-seq) consiste à séquencer l'ARN au sein de cellules individuelles, permettant ainsi d'identifier des types cellulaires distincts et d'étudier les trajectoires de différenciation cellulaire. Cette technologie révolutionne des domaines comme la biologie du développement, l'immunologie et la recherche sur le cancer, en offrant une résolution sans précédent des profils transcriptionnels.

Bien que les méthodes d'analyse ne soient pas encore entièrement standardisées, les principales étapes du pipeline comprennent :

- Contrôle qualité et filtrage des données: élimination des cellules présentant un faible nombre de gènes détectés, ou un pourcentage élevé de transcrits mitochondriaux.
- Normalisation : ajustement des comptages pour corriger les différences de profondeur de séquençage entre les cellules.
- Sélection des gènes hautement variables (HVG) : identification des gènes qui contribuent le plus à la variabilité entre les cellules.
- Réduction de dimensionnalité (ACP): simplification des données pour mieux visualiser les structures complexes.
- Clustering : regroupement des cellules en clusters, correspondant à des types cellulaires.
- Annotation fonctionnelle: interprétation des clusters en fonction de marqueurs connus ou d'enrichissements fonctionnels.

En plus de ces étapes classiques l'analyse de scRNA-seq offre d'autres possibilités qui permettent une compréhension plus complète des systèmes biologiques avec par exemple la transcriptomique spatiale ou encore le pseudo time analysis.

Au cours de ce projet nous avons réalisé l'analyse de données scRNA-seq obtenues dans le cadre d'une étude sur les effets de l'exercice physique sur l'évolution des populations cellulaires et l'expression des gènes des cellules du tissu musculaire. Alen Lovrić et son équipe ont publié l'article "Single-cell sequencing deconvolutes cellular responses to exercise in human skeletal muscle" en 2022 dans la revue Nature, dans lequel ils s'intéressent à la plasticité du tissu musculaire en réponse à l'exercice. Le bon fonctionnement des muscles et la masse musculaire squelettique sont très importants dans la santé, et ont un impact sans équivalent sur la qualité de vie et les maladies, particulièrement au cours du vieillissement. Bien que les processus d'adaptation musculaire à l'exercice aient été étudiés au niveau physiologique (au niveau de l'organe musculaire), de nombreuses questions demeurent ouvertes quant à l'adaptation au niveau cellulaire voir génétique. Dans cette étude, ils ont utilisé le scRNA-seq pour caractériser des cellules individuelles issues de biopsies de muscles squelettiques prélevées avant et après un exercice physique intense. L'objectif principal était de caractériser la composition cellulaire du muscle squelettique adulte au repos et de déconvoluer les réponses transcriptionnelles à l'exercice au sein des différentes populations cellulaires.

Durant cette étude nous allons présenter trois fonctionnalités optionnelles qui peuvent (ou doivent dans certains cas) être implémentées au cours de l'analyse de données scRNA-seq pour aller plus loin. En effet en plus du pipeline classique nous avons réalisé l'intégration de plusieurs échantillons grâce au principe d'anchoring, nous avons également réalisé une analyse d'expression différentielle entre les conditions pré exercice et post exercice dans tous les types cellulaires en implémentant le tutoriel differential expression testing du site satijalab.org/seurat et nous avons finalement réalisé une analyse de gene ontology pour comprendre les voies cellulaires impliquées dans la réponse à l'exercice avec le package ClusterProfiler.

Matériel et Méthode

Protocole d'activité physique :

L'étude a inclus trois individus en bonne santé (une femme et deux hommes) âgés de 18 à 50 ans. Les participants avaient une activité physique régulière (au moins deux fois par semaine), sans être des athlètes professionnels. Les critères d'exclusion incluent la prise de médicaments actifs, la participation à des sports organisés à un niveau athlétique, ainsi que l'usage régulier de produits à base de tabac.

Chaque participant a subi deux biopsies musculaires du vaste latéral permettant de prélever environ 1 g de tissu par prélèvement. Les biopsies ont été réalisées immédiatement avant et 3 heures après l'exercice physique. L'exercice consistait en trois sprints de 30 secondes sur un ergomètre à freinage mécanique, séparés par 2 minutes de récupération active.

Accessibilité des données :

Les données de séquençage générées dans cette étude sont publiquement accessibles via la plateforme Gene Expression Omnibus (GEO) sous le numéro d'accèsion GSE214544. Les données utilisées sont constituées de 6 fichiers (un par sujet avant exercice et un par sujet après exercice) contenant chacun un fichier barcode.tsv, features.tsv et matrix.mtx. L'analyse a été réalisée en R avec le package *seurat* et les données ont été chargées sous forme de matrice sparse. Les objets *seurat* ont alors été créés pour chaque échantillon et pour chaque condition en conservant uniquement les gènes exprimés dans au moins 5 cellules et les cellules exprimant au moins 200 gènes pour une première filtration des données de mauvaise qualité. En condition pré exercice on obtient pour l'individu 1 : 13614 gènes pour 2425 cellules, pour l'individu 2 : 14861 gènes pour 4769 cellules et pour l'individu 3 : 15621 gènes pour 15516 cellules. En condition post exercice on obtient pour l'individu 1 : 13614 gènes pour 3057 cellules, pour l'individu 2 : 14390 gènes pour 1435 cellules et pour l'individu 3 : 14826 gènes pour 9566 cellules.

Résultats

Filtering des données

Notre première véritable étape de pré-processing a été la suppression des cellules de mauvaise qualité correspondant à des cellules sous séquençé, des débris cellulaires ou bien des morceaux d'ARN flottant dans la solution, par une analyse qualitative (nombre de gènes et d'UMI) dans chacun des échantillons.

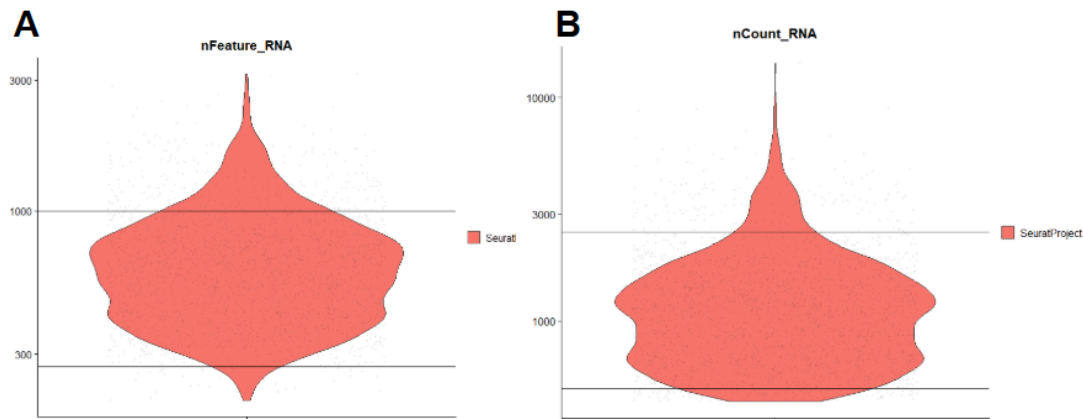


Figure.1 : Violin plot du nombre gène (A) et d'UMI (B) pour le sujet 1 avant exercice

Par exemple dans le cas de l'échantillon pré exercice du premier participant il a été décidé de conserver les cellules exprimant un nombre de gènes compris entre 270 et 1000 et ayant un nombre d'UMI compris entre 500 et 2500 (figure 1).

	Nombres de gènes	Nombre d'UMI
Sujet 1 avant exercice	270 - 1000	500 - 2500
Sujet 2 avant exercice	270 - 1000	500 - 2500
Sujet 3 avant exercice	270 - 1000	400 - 1000
Sujet 1 après exercice	270 - 1000	500 - 2500
Sujet 2 après exercice	500 - 3000	100 - 8000
Sujet 3 après exercice	270 - 1000	400 - 1500

Tableau.1 : Tableau récapitulatif des seuils pour le nombre de gène et d'UMI pour les cellules retenues dans la suite de l'analyse pour chacun des 6 échantillons

Les cellules de bonne qualité ont été retenue grâce à la fonction suivante pour chaque échantillon : `subset([nom de l'échantillon], subset = nFeature_RNA > minGene & nFeature_RNA < maxGene & nCount_RNA > minUMI & nCount_RNA < maxUMI)` (cf seuils tableau 1).

Ensuite nous avons filtré les cellules ayant un taux d'ARN mitochondrial supérieur à 15% comme dans l'article de référence. En effet les cellules avec un taux important d'ARN mitochondrial sont des cellules stressées, sénescents voir mortes qui sont donc de mauvaise qualité et non pertinentes pour nos analyses.

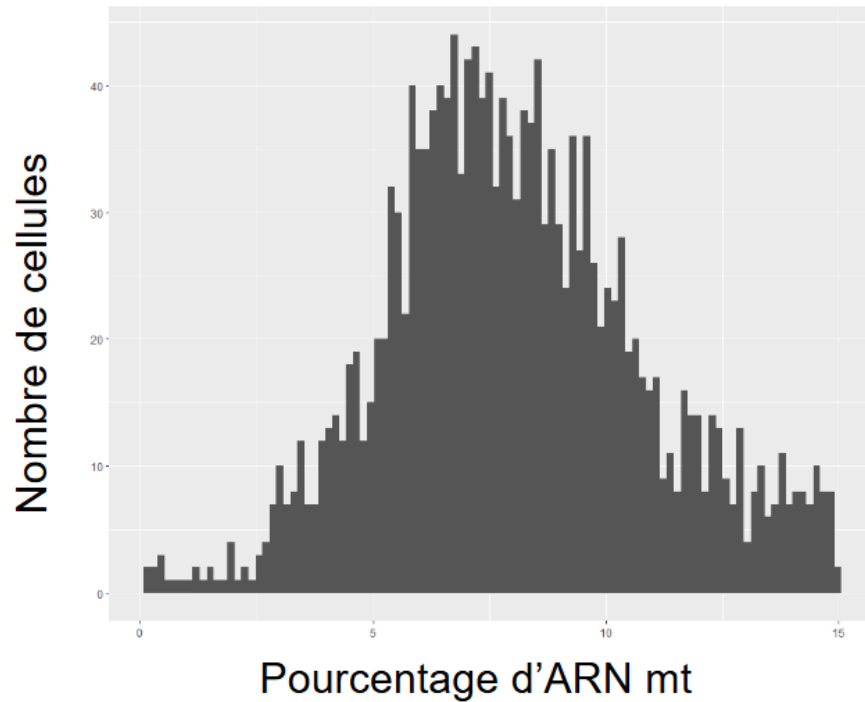


Figure.2 : Pourcentage d'ARN mitochondrial après filtrage pour le sujet 1 avant exercice

Nous avons filtré ces cellules grâce à la fonction `PercentageFeatureSet()` qui nous permet de calculer le pourcentage d'ARN mitochondrial par cellules en renseignant le pattern des gènes mitochondriaux, '^MT-' dans notre cas. Les cellules ayant un taux d'ARN mitochondrial inférieur à 15% sont ensuite sélectionnées via la commande `subset()` sur le paramètre `percent_mt` pour chaque échantillon (figure 2).

Elimination des doublets

Ensuite nous avons supprimé les doublets et les multiplets provenant du fait que deux cellules ou plus soit entrée dans une même droplet lors de l'expérimentation conduisant à des cellules ayant un profil d'expression aberrant et non pertinent. Pour cela nous avons choisi d'utiliser `ScDblFinder()`, ainsi nous avons tout d'abord converti notre objet Seurat en objet single-cell experiment via `as.SingleCellExperiment()` puis appliqué la fonction `ScDblFinder()` qui génère des doublets artificiels et applique un modèle des k plus proches voisins (knn) pour identifier les doublets. Pour finir nous avons sélectionné uniquement les singlets avec `subset()` sur le paramètre `doublets.class`.

	Nombres de gènes	Nombre de cellule
Sujet 1 avant exercice	1721	12463
Sujet 2 avant exercice	3384	13812
Sujet 3 avant exercice	9330	14316
Sujet 1 après exercice	2191	12470
Sujet 2 après exercice	1173	13713
Sujet 3 après exercice	6932	13891

Tableau.2 : Nombre de cellules et de gènes finalement conservés après les différentes étapes de pré-processing pour les 6 échantillons.

Normalisation par SCTransform

Nous avons procédé à la normalisation des données sur les 6 échantillons. Certaines cellules peuvent être sous séquencé (biais technique) ou certaines cellules expriment plus ou moins d'ARN naturellement. Pour pallier à ces phénomènes on effectue une normalisation, nous avons choisi d'utiliser la fonction `SCTransform()` qui se base sur les modèles linéaires généralisés. Nous avons précisé à la fonction de régresser pour `nCount_RNA` et `percent_mt`. La fonction `SCTransform()` s'occupe également de l'étape de sélection des gènes hautement variables et de l'étape de scaling, autrement dit chaque expression de gènes est divisée par sa moyenne ainsi les gènes les plus exprimé n'ont pas plus d'importance que les autres dans les étapes suivantes du pipeline.

PCA

Une fois ces étapes terminées, nous avons procédé à l'analyse en composantes principales (PCA). Bien que nous ayons initialement sélectionné 2 000 gènes hautement variables (HVGs), ce nombre reste trop élevé pour les analyses en aval, rendant nécessaire une réduction de dimensionnalité. Nous avons effectué la PCA en utilisant la fonction `RunPCA()` sur nos objets Seurat filtrés, normalisés et mis à l'échelle des différents échantillons.

Intégration

Avant l'intégration, nous avons veillé à spécifier les identifiants d'origine (`orig.ident`) pour pouvoir tracer les échantillons après l'intégration.

Une fois les étapes de prétraitement réalisées pour les 6 échantillons, nous sommes passés à l'intégration des données. L'objectif principal de cette étape est de corriger les batch effects, c'est-à-dire les variations techniques entre échantillons, tout en préservant les signaux biologiques pertinents.

Pour cela nous avons utilisé la fonction `SelectIntegrationFeatures()`, cette dernière fonctionne comme `FindVariableFeatures()` sauf qu'elle permet de faire cette étape nos 6 échantillons en même temps. Cela permet d'identifier les gènes hautement variables robustes à travers les échantillons, dans les faits cette fonction donne un score en fonction du nombre d'échantillon dans lequel le gène est un hvg (6/6 si il vari dans nos 6 échantillons, 0 si il ne vari dans aucun échantillon), pour les gènes ayant le même score la fonction va regarder le score de la features qui varie dans chaque échantillon pour départager. Ici nous avons choisi de garder 2000 gènes variables donc la fonction a renvoyé la liste des 2000 gènes qui varient le plus dans les 6 sets.

Nous avons ensuite utilisé la fonction `FindIntegrationAnchors()`, cette dernière permet de trouver les ancrs, c'est-à-dire les paires de cellules des différents échantillons qui sont plus proche voisines mutuelles lors de la réduction produite par cette fonction, les ancrs de faible confiance sont ensuite supprimées en comparant leur pertinence par rapport à l'espace de grande dimension. Enfin les ancrs restantes sont utilisées via la fonction `IntegrateData()` pour aligner les valeurs des 6 échantillons en un seul que nous avons utilisé pour procéder à notre clustering.

Ensuite nous avons de nouveau mis à l'échelle via `ScaleData()` et réduit la dimension des données intégrées via `RunPCA()` avant de procéder au clustering.

Clustering et annotation

Tout d'abord, nous créons le graphe s-NN (graphe des voisins partagés) à l'aide de la méthode `FindNeighbors`, puis nous le partitionnons avec la méthode `FindClusters`, qui est basée sur l'algorithme de Louvain. Après plusieurs itérations sur le nombre d'axe de PCA et la résolution lors de la partition du graph, nous avons déterminé que le

meilleur clustering était produit avec un nombre d'axe de 15 et une résolution de 0,2. Avec ces paramètres on obtient alors 8 clusters qui ont été annoté par annotation manuelle en utilisant une liste des marqueurs de notre article de référence. Par exemple pour les cellules endothélial nous avons regardé le facteur de Von Willebrand (VMF) connue pour son rôle dans le recrutement des plaquettes au site de dommage vasculaire et qui est principalement exprimé par les cellules endothéliales¹ ou bien ESAM qui est la protéine d'adhésion typique des cellules endothélial (figure 3.A). Ainsi les cellules qui expriment fortement les marqueurs d'un type cellulaire et uniquement ceux-là sont annoté comme étant de ce type cellulaire (exemple avec les cellules des clusters 1 et 6 qui expriment très fortement et uniquement les marqueurs des cellules mésenchymateuses). La liste détaillée des marqueurs pour chaque type cellulaire (couleurs correspondants au cluster) sont présentés en annexe 1. Lors de ce clustering un groupe de cellules n'a pas été identifié avec les marqueurs que nous avions à notre disposition.

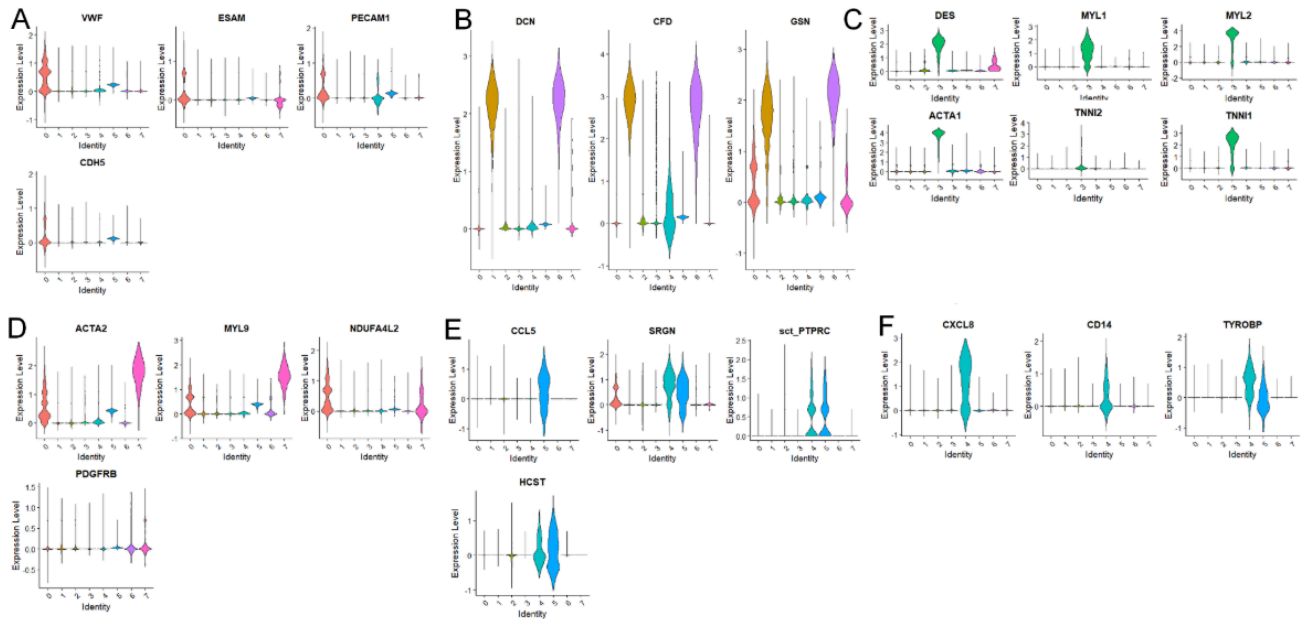


Figure.3 Violin plot pour les marqueurs exprimé dans les différents clusters

- A) Expression des marqueurs des cellules endothéliales
- B) Expression des marqueurs des cellules mésenchymateuses
- C) Expression des marqueurs des cellules myogéniques
- D) Expression des marqueurs des péricytes
- E) Expression des marqueurs des lymphocytes
- F) Expression des marqueurs des monocytes

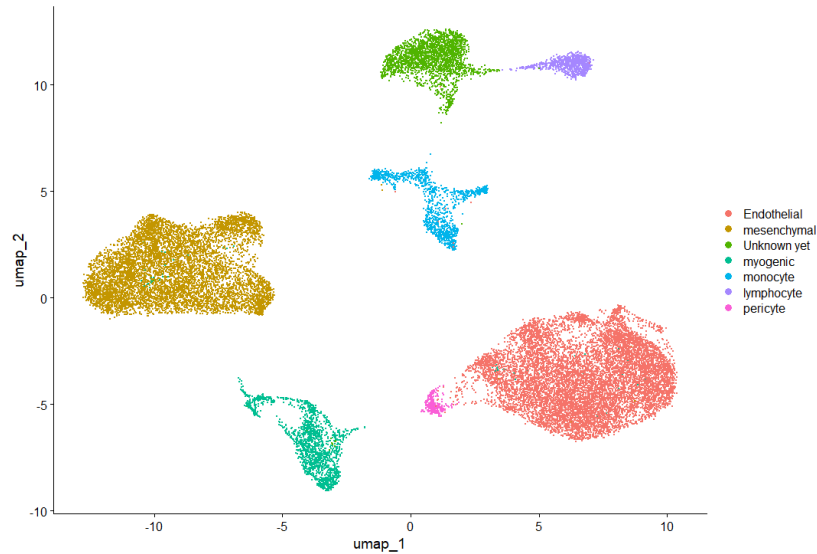


Figure.4 UMAP obtenu après clustering (PC =15, résolution = 0.2)

On obtient ainsi une UMAP avec nos clusters annoté (figure 4), la proportion des différents types cellulaires est similaire à celle de l'article.

Evolution de la proportion des types cellulaires en réponse à l'exercice

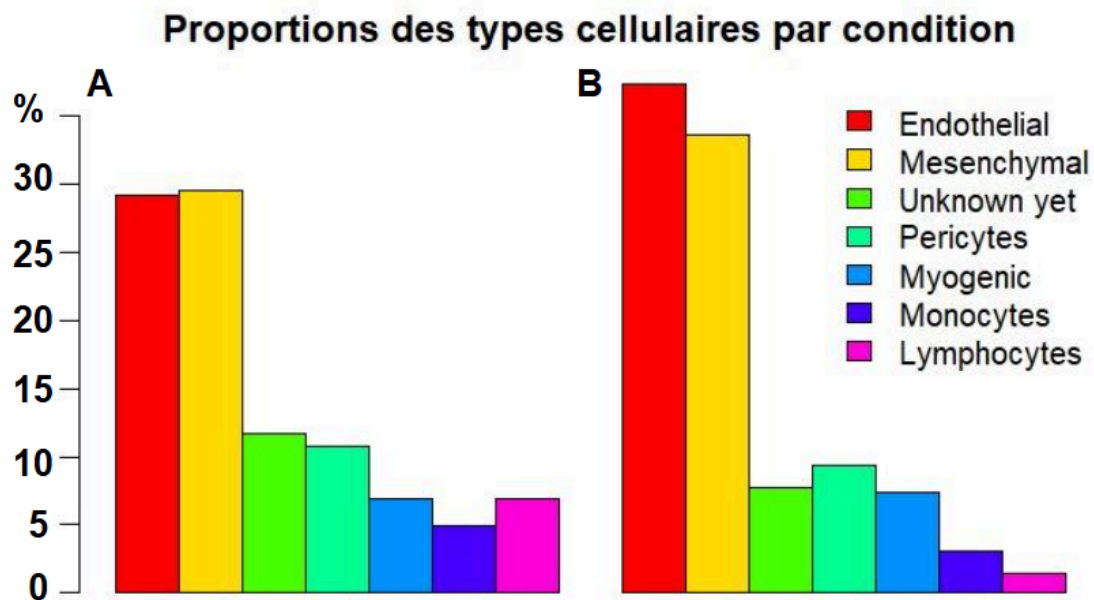


Figure.5 Composition cellulaire après (A) et avant exercice(B).

On observe ici que la proportion de lymphocyte augmente après l'effort passant de 1.4% à 6.4%, il en est de même pour la proportion de monocyte qui augmente de 5.3% (3,1% -> 8.4%) ce qui est concordant avec les observations de notre article de référence (figure 5). Ce phénomène est bien connu dans la littérature², on peut faire l'hypothèse

que l'activation du système nerveux sympathique en réponse au stress induit par l'activité physique conduit à la libération de catécholamine dans le sang ce qui conduit au recrutement des lymphocytes et des monocytes d'où le changement de proportion cellulaire. Parallèlement on observe une diminution de la proportion de cellules endothéliales (45.4% -> 36.3%), de cellules mésenchymateuses (33.5% -> 28.0%) et de péricyte (12.7% -> 10.3%) mais plus qu'une diminution réelle, il s'agit de la conséquence de l'augmentation de la proportion de cellules immunitaires (figure 5). La proportion de cellules myogéniques reste elle inchangée. Dans le cluster que nous n'avons pas réussi à identifier on observe une augmentation de la proportion cellulaire (7.7% -> 12.4%), on peut donc supposer qu'il s'agit d'un autre type de cellule immunitaire ou bien qu'une partie importante de ce cluster est constitué de lymphocyte et de monocyte que nous n'avons pas réussi à bien déterminer (figure 5).

Expression différentielle en réponse à l'exercice et gene ontology

Enfin, une analyse d'expression différentielle des gènes en réponse à l'exercice a été réalisée et ce pour chacun de 6 types cellulaires identifiés. Pour ce faire nous avons rassemblé sous la condition pré les cellules des trois sujets provenant des échantillons prélevés avant l'exercice et sous la condition post les cellules provenant des échantillons obtenu après l'exercice pour les 3 sujets. En suivant le tutoriel Differential expression testing proposé par Satija lab (satijalab.org/seurat/articles/de_vignette) nous avons pu réaliser l'analyse d'expression différentielle en utilisant la fonction FindMarkers. En effet, ce tutoriel présente un test basé, comme dans l'article, sur le test non paramétrique Wilcoxon pour tester les gènes différentiellement exprimés entre deux groupes spécifiques de cellules, ici les cellules pré et post exercice.

Le test d'expression différentielle effectué avec la fonction FindMarkers de Seurat génère une liste de gènes significativement modulés entre deux groupes comparés, accompagnée de plusieurs métriques. Parmi celles-ci, on retrouve la p-value, qui évalue la significativité statistique de la différence d'expression, et le log fold change (logFC), qui indique l'ampleur et la direction de cette différence. Une p-value ajustée (p_adj) est également fournie pour corriger les tests multiples, garantissant une interprétation robuste des résultats. Dans notre cas on s'intéressait aux gènes surexprimés en réponse à l'exercice physique, on a donc fait une sélection des gènes ayant une p-value ajustée < 0.05 et un log fold change > 0.5 dans chaque type cellulaire. Une fois les gènes surexprimés en réponse à l'exercice identifiés, une gene ontology a été réalisée dans l'optique d'étudier les voies et les processus biologiques les plus importants dans la réponse à l'exercice dans les différents types cellulaires. Pour ce faire le package ClusterProfiler a été utilisé.

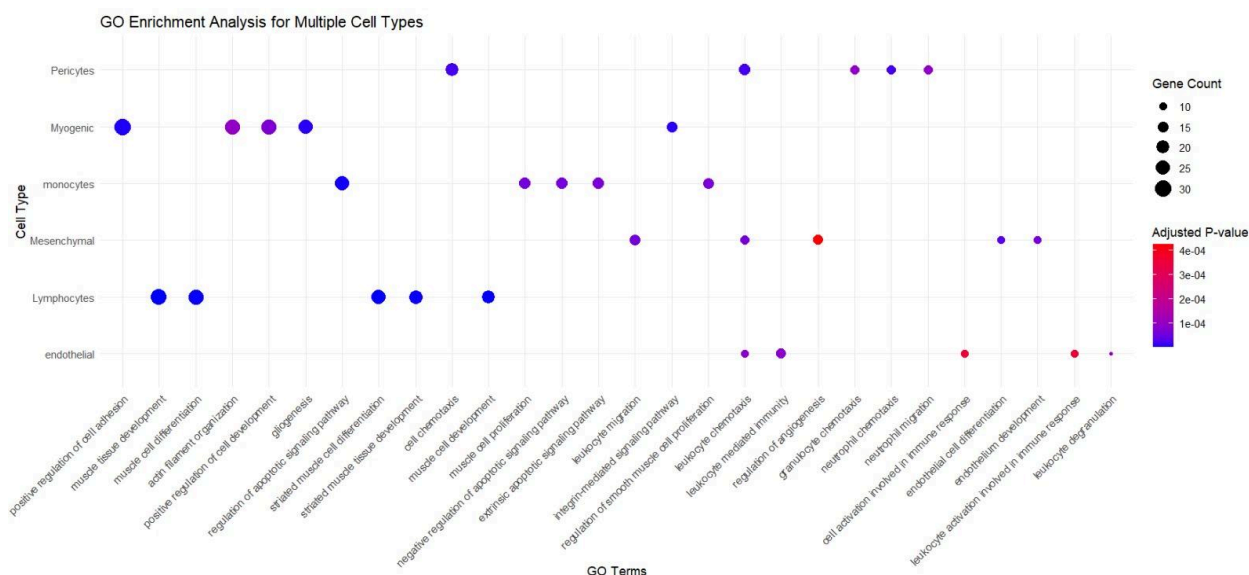


Figure.6 : Gene ontology sur les gènes répondant le plus à l'exercice physique dans les différents types cellulaires.

On observe à travers tous les types cellulaires 1440 gènes (incluant les répétitions de gènes d'un type cellulaire à l'autre) différenciellement exprimés.

Les cellules myogéniques sont celles qui présentent le plus grand nombre de gènes différenciellement exprimés entre la condition pré et post exercice. L'analyse de gene ontology révèle que les gènes surexprimés dans la condition post sont impliqués dans l'organisation de la matrice extracellulaire et des filaments d'actine (eux même impliqués dans la contraction musculaire) avec des fonctions comme "actin filament organisation" (figure 6). Il est tout à fait cohérent de retrouver dans les cellules myogéniques une surexpression des gènes qui codent pour une structure impliquée dans la contraction musculaire, et c'est d'ailleurs ce qui a été également observé dans l'article au cours de l'analyse originelle. On retrouve également une implication dans la gliogenèse en réponse à l'exercice (figure 6). En comparaison à l'article d'origine on ne retrouve pas précisément les voies de réponse au stress comme "negative regulation of cell death" et "regulation of growth" parmi les voies cellulaires principales mais on a également dans notre analyse des processus qui vont dans le même sens, comme "positive regulation of cell development" (figure 6).

Ensuite on a les lymphocytes avec 322 gènes différenciellement exprimés. La gene ontology révèle, pour ce type cellulaire, que les gènes surexprimés sont impliqués dans le développement musculaire et la différenciation des cellules musculaires avec des voies tels que "muscle tissue development", "muscle cell differentiation", "striated muscle cell differentiation" ou encore "striated muscle tissue development" (figure 6). On a donc un impact des lymphocytes en réponse à l'exercice qui semble pointer clairement vers le développement musculaire. Dans l'article ils semblent n'avoir trouvé pour les lymphocytes un enrichissement que pour des termes génériques ce qui suggère une absence de réponse spécifique à l'exercice des lymphocytes.

Concernant les péricytes on observe 322 gènes différenciellement exprimés et principalement associés au chimiotactisme donc le déplacement dirigé ou la migration (ici vers le tissu musculaire) des cellules immunitaires avec des voies cellulaires tels que "leukocyte chemotaxis", "neutrophil chemotaxis" ou "neutrophil migration" (figure 6). Dans le muscle, ces processus pourraient refléter une réponse inflammatoire ou réparatrice induite par l'exercice, en particulier après des efforts intenses provoquant des micro-lésions musculaires. On ne retrouve pas directement les mêmes voies que dans l'article où les fonctions sont plutôt liées à l'activation cellulaire, la régénération et la réponse au stress, cependant le recrutement de cellules immunitaires peut être également une réponse au stress et être impliqué dans la régénération.

Les monocytes présentent 268 gènes différenciellement exprimés lesquels sont impliqués dans les fonctions de régulation de l'apoptose et le développement musculaire (figure 6). Les auteurs de l'article n'ont pas trouvé pour ce type cellulaire de fonction spécifique de réponse à l'exercice.

Les cellules mésenchymateuses montrent une expression différentielle pour 143 gènes et ces derniers sont principalement impliqués dans la régulation de l'angiogenèse, le développement et la différenciation de l'endothélium et, tout comme les péricytes, la migration et le chimiotactisme des cellules immunitaires (figure 6). L'angiogenèse en réponse à l'exercice physique est un phénomène bien connu³ il s'agit d'un processus important par lequel de nouveaux vaisseaux sanguins se forment dans les muscles en réponse à l'activité physique. Concernant le développement de l'endothélium, il a déjà été montré que l'exercice physique avait un impact sur la fonction et le développement de l'endothélium⁴ notamment via l'augmentation du débit sanguin ce qui va dans le même sens que les effets sur l'angiogenèse. Dans l'article, ils ont trouvé des fonctions telles que la régénération et le remodelage des organes ou la guérison des blessures, ce qui va dans le même sens que ce qui a été trouvé dans cette analyse.

Enfin on observe seulement 90 gènes différenciellement exprimés en réponse à l'exercice dans les cellules endothéliales et ces gènes sont principalement impliqués dans des fonctions d'activation et de chimiotactisme des cellules immunitaires avec des voies tels que "leukocyte chemotaxis" et "leukocytes activation involved in immune

response” (figure 6). Dans l'article, ils ont également retrouvé des caractéristiques ontologiques liées à la réponse au stress et à l'activation cellulaire.

Discussion

Ce projet a donc consisté à réaliser une nouvelle analyse des données de l'article présenté plus haut et intégré de nouvelles pratiques qui n'ont pas été vu en cours. Nous avons commencé par réaliser un pré-processing classique des données des différents échantillons puis nous avons normalisé les données et réalisé une réduction de dimension par PCA. Les données ont ensuite été intégrées puis un clustering et une annotation fonctionnelle ont été effectués. Une fois les types cellulaires obtenus, les effets de l'exercice sur la composition cellulaire et l'expression des gènes des cellules du tissu musculaire ont pu être analysés.

Il en ressort que l'exercice physique modifie la composition cellulaire en augmentant la proportion de lymphocytes et monocytes, deux types de cellules immunitaires retrouvés au niveau du tissu musculaire, et diminue en conséquence la proportion des autres types cellulaires.

Cette augmentation du nombre de cellules immunitaire en réponse à l'exercice peut être expliqué par la migration des cellules immunitaires induite par certains types cellulaires comme on le voit dans la gene ontology avec les cellules endothéliales, mésenchymateuse et les péricytes chez lesquels ont été observé une forte expression de gènes impliqués dans le chimiotactisme des cellules immunitaires. Cette augmentation des cellules immunitaires a déjà été observée dans la littérature². Ces phénomènes n'ont pas été retrouvés dans l'analyse en gene ontology des auteurs de l'article, bien qu'ils observent également une augmentation de la proportion de lymphocyte et monocyte après l'exercice.

On retrouve également pour les cellules myogéniques deux résultats de gene ontology qui n'ont pas été présentés dans l'article. Tout d'abord il apparaît que les cellules myogéniques surexpriment des gènes impliqués dans la voie de signalisation médiée par l'intégrine. Cela est cohérent avec des résultats de la littérature qui ont démontré que l'ARNm et la protéine de la sous-unité $\alpha 7$ -intégrine sont régulés à la hausse après des contractions excentriques, comme un mécanisme visant à renforcer les structures supportant la charge et à résister aux blessures lors de répétitions d'exercices. Il est également hypothétisé depuis de nombreuses années que l'intégrine peut favoriser le renouvellement des protéines d'une manière qui soutient les adaptations bénéfiques à l'entraînement en résistance, y compris l'hypertrophie⁶.

Les cellules myogéniques surexpriment également des gènes impliqués dans la gliogenèse, la synthèse de cellules gliales, ce qui ne semble pas avoir de sens en réponse à l'exercice au premier abord, cependant l'exercice physique peut induire des adaptations nerveuses, notamment une meilleure efficacité de la transmission neuromusculaire. La gliogenèse pourrait alors refléter une réponse adaptative liée à la myélinisation ou au soutien des motoneurones par les cellules gliales, favorisant une propagation plus rapide et efficace du message nerveux vers les muscles. Ces adaptations contribuent à l'amélioration de la coordination et de la performance musculaire.

Les quelques différences observées entre nos résultats et ceux de l'article peuvent s'expliquer par plusieurs facteurs. Bien qu'ils aient également utilisé une méthode d'analyse d'expression différentielle basée sur le test de rang de Wilcoxon, les détails précis de leur approche ne sont pas disponibles d'autant plus qu'une partie a été réalisée avec Scanpy contrairement à nous. Il est donc possible qu'il y ait des variations dans la manière dont ils ont effectué leurs tests, notamment pour la Gene Ontology, où l'on ignore s'ils ont utilisé clusterProfiler. En fin de compte, ces différences de résultats peuvent s'expliquer par le fait que nous n'avons pas suivi exactement les mêmes processus ni appliqué les mêmes seuils à chaque étape.

En conclusion on retrouve des résultats logiques, pour la plupart similaires avec ceux de l'article et ce qui n'ont pas été présenté dans l'article sont cohérent par rapport à la littérature.

Références

- 1)Nightingale, T., and D. Cutler. ‘The Secretion of von Willebrand Factor from Endothelial Cells; an Increasingly Complicated Story’. *Journal of Thrombosis and Haemostasis*, vol. 11, no. Suppl 1, June 2013, pp. 192–201. *PubMed Central*, <https://doi.org/10.1111/jth.12225>.
- 2)Neves, Pedro Rogério Da Silva, et al. ‘Acute Effects of High- and Low-Intensity Exercise Bouts on Leukocyte Counts’. *Journal of Exercise Science and Fitness*, vol. 13, no. 1, June 2015, pp. 24–28. *PubMed Central*, <https://doi.org/10.1016/j.jesf.2014.11.003>.
- 3)Ross, Mark, et al. ‘Exercise-Induced Skeletal Muscle Angiogenesis: Impact of Age, Sex, Angiocrines and Cellular Mediators’. *European Journal of Applied Physiology*, vol. 123, no. 7, July 2023, pp. 1415–32. *Springer Link*, <https://doi.org/10.1007/s00421-022-05128-6>.
- 4)Di Francescomarino, Samanta, et al. ‘The Effect of Physical Exercise on Endothelial Function’. *Sports Medicine (Auckland, N.Z.)*, vol. 39, no. 10, 2009, pp. 797–812. *PubMed*, <https://doi.org/10.2165/11317750-000000000-00000>.
- 5)Boppart, Marni D., and Ziad S. Mahmassani. ‘Integrin Signaling: Linking Mechanical Stimulation to Skeletal Muscle Hypertrophy’. *American Journal of Physiology - Cell Physiology*, vol. 317, no. 4, Oct. 2019, pp. C629–41. *PubMed Central*, <https://doi.org/10.1152/ajpcell.00009.2019>.
- 6)Takaya, Kento, et al. ‘Decorin Inhibits Dermal Mesenchymal Cell Migration and Induces Scar Formation’. *Plastic and Reconstructive Surgery. Global Open*, vol. 10, no. 4, Apr. 2022, p. e4245. *PubMed*, <https://doi.org/10.1097/GOX.0000000000004245>.

Annexes

VMF	facteur de Von Willebrand (hémostasie)
ESAM	Protéine de liaison des cellules endothélial
PECAM1	Protéine d'adhésion aux cellules endothélial et aux plaquettes 1
CDH5	Cadhérine vasculaire endothélial
DCN	Décorine limitant la migration des cellules mésenchymateuse ⁶
CFD	Facteur de complément D
GSN	Gelsoline (transition cellulaire épithélial -> mésenchymateuse)
DES	Desmine marqueur myogénique connu
MYL1,2	Chaîne légère de la myosine
ACTA1	Actine alpha 1, muscle spécifique
PAX7	Marqueur des cellules satellites musculaires
TNNI1,2	Troponine 1 et 2
ACTA2, PDGFRB	Marqueurs connus des péricytes
MYL9	Polypeptide de régulation 9 de la chaîne légère de la myosine
NDUFA4L2	Complexe mitochondrial spécifique aux péricytes
IGFBP7	Facteurs de croissance chez les péricytes et les cellules endothéliales
CCL5	Exprimé chez les LT activés
SRGN	Protéoglycane intracellulaire chez les cellules d'origine hématopoïétique
PTPRC	Tyrosine phosphatase transmembranaire chez les LT
B2M	Beta-2-microglobuline, essentiel complex d'histocompatibilité chez les CD4+
HCST	Protéine de transduction chez les cellules hématopoïétique
CXCL8	IL-8 qui stimule la migration des monocytes
CD14	Récepteur typique des monocytes
TYROBP	Protéine de liaisons à la tyrosine exprimé chez les monocytes périphérique au système sanguin
FTL	Ferritine exprimé par les monocytes périphérique au système sanguin

Annexe 1 : Liste des marqueurs utilisés pour l'annotation du clustering.